Jia, Y., Andrieu, C., Piechocki, RJ., & Sandell, M. (2008). Depth-first and breadth-first search based multilevel SGA algorithms for near optimal symbol detection in MIMO systems. *IEEE Transactions on Wireless Communications*, *7*(3), 1052 - 1061. https://doi.org/10.1109/TWC.2008.060813

Depth-First and Breadth-First Search Based Multilevel SGA Algorithms for Near Optimal Symbol Detection in MIMO Systems

Yugang Jia, *Student Member, IEEE,* Christophe Andrieu, Robert J. Piechocki, *Member, IEEE,* and Magnus Sandell, *Member, IEEE*

***Abstract*— The multilevel structure of the N-QAM modulation constellations is exploited to significantly reduce the complexity of the sequential Gaussian approximation (SGA) algorithm [1] for near optimal symbol detection in spatial multiplexing multiple-input multiple-output (MIMO) system. We propose two multilevel SGA algorithms (MSGA) which are based on depth-first search (DFS) and breadth-first search (BFS) respectively. Additionally, an important methodological contribution to this multilevel technique is proposed where the mismatch between the pseudo symbols and the true symbols is taken into consideration for the computation of posterior probabilities of symbol combinations. We justify this from a theoretical perspective as well as with numerical results. Simulation results show that the performance of the two proposed multilevel algorithms can approach that of the optimal a posteriori probability (APP) detector while its total computation cost is at most 81% and 48% of that of the original SGA algorithm for 16QAM and 64QAM modulation MIMO systems with 4 transmit/receive antennas respectively.**

***Index Terms*— Complexity reduction, Gaussian approximation, multilevel modulation, multiple-input multiple-output (MIMO) systems.**

I. Introduction

THE use of multiple-input multiple-output (MIMO) architectures [2] promises to achieve high capacity for wireless communication channels in rich multipath environments. High order QAM constellations are usually adopted to improve spectral efficiency in such systems, which makes it difficult to use maximum likelihood (ML) detection due to its intractable complexity.

Computational efficient symbol detection algorithms have been widely explored to achieve the substantial performance gains promised by spatial multiplexing MIMO systems with QAM constellations. The various sphere decoders (SD) [3] [4] [5] [6] tend to approach the optimal performance efficiently but suffer from the fact that their complexity is channel and SNR dependent [7]. Other approaches include algorithms based on the Gaussian approximation principle, a.k.a. probabilistic data association (PDA) [8] [9] [10], but these results do not carry on to high order modulations (16QAM/64QAM).

The sequential Gaussian approximation (SGA) algorithm [1] has been demonstrated to achieve near optimal performance with fixed complexity and memory requirement. The key step of the SGA algorithm consists of sequentially identifying a reduced number M of highly probable symbol combinations for antennas $1, \ldots, j$ with $j = 1, \ldots, N_T$. In each step, only the M significant symbol combinations are selected via evaluating the likelihoods of all MN possibilities (N is the number of symbols in modulation alphabet A) and kept for the next step, until the N_T-th antenna is reached. Then, the M significant symbol combinations for all the antennas are used in order to compute the marginal posterior probabilities. This results in a significant complexity reduction and very good performance has been observed in computer simulations. Although the complexity of the SGA algorithm is less than that of the SD [12], it does not lend itself to an efficient implementation for MIMO systems with large constellation size, in particular due to the evaluation and sorting of the likelihoods involved in the algorithm.

Fortunately, large QAM constellations exhibit a natural multilevel structure. The N-QAM constellations can be decomposed into $L \stackrel{\text{def}}{=} \log_4(N)$ levels [1] where in each level a set of pseudosymbols can be constructed from pseudosymbols set in a lower level.

The multilevel structure of the N-QAM constellation has been widely exploited in the literature for complexity reduction purpose. In [13] [14] [15], an iterative tree search (ITS) algorithm is proposed for turbo detection of MIMO systems. The ITS scheme is based on a reduced search space via the use of the **M** algorithm [17] in conjunction with the use of multilevel bit mappings. It is also shown that the complexity of ITS per bit is only dependent on the the length of information blocks and independent of the constellation size N. In [16], a multilevel sampling scheme is proposed to reduce the complexity of the mixture Kalman filter for adaptive detection of 16-QAM symbols over flat-fading channels. The simulation results show that the proposed multilevel mixture Kalman filter achieves a performance similar to that of the original mixture

Manuscript received October 11, 2006; revised February 15, 2007 and May 24, 2007; accepted August 20, 2007. The associate editor coordinating the review of this paper and approving it for publication was F. Daneshgaran. This work was supported by Toshiba Research Europe Ltd (Bristol), UK.

Y. Jia was with University of Bristol, Bristol, UK. He is now with Philips Research Asia, Shanghai, P.R. China (e-mail: yugang.jia@philips.com).

R. J. Piechocki is with the Centre for Communications Research, University of Bristol, Bristol, BS8 1UB, UK (e-mail: c.andrieu@bristol.ac.uk).

C. Andrieu is with the Department of Mathematics, University of Bristol, Bristol, BS8 1TW, UK (e-mail: r.j.piechocki@bristol.ac.uk).

M. Sandell is with Toshiba Telecommunication Research Lab, Bristol, BS1 4ND, UK (e-mail: magnus.sandell@toshiba-trel.com).

Digital Object Identifier 10.1109/TWC.2008.060813.

[1] L can only be an integer.

Kalman filter, but with a much lower complexity.

In this paper, we propose two reduced complexity SGA algorithms that exploit the natural hierarchical approximating structure which has been suggested for QAM constellations. The first proposed multilevel scheme is based on a depth-first search (DFS) which has been proposed previously in the literature [13] [14] [15] [16], but not in the context of the SGA algorithm. The second one is based on the breadth-first search (BFS) which has not been suggested in the multilevel literature. Both algorithms (MSGA-DFS and MSGA-BFS) aim to select a set of M most significant symbol combinations for computation of the posterior marginal symbol probabilities. The complexity burden of computation and sorting of likelihoods involved in the MSGA algorithms is only $1/2$ and $3/16$ of that of the SGA algorithm for MIMO systems with 16QAM and 64QAM constellations respectively. The exact complexity reduction will be explained later.

Additionally, in the course of this research, we have made an important methodological contribution to the multilevel literature, for which we have developed a theoretical justification. For both MSGA algorithms (MSGA-DFS and MSGA-BFS), the mismatch between the pseudosymbols and the true symbols is taken into consideration for the computation of posterior probabilities of symbol combinations. More specifically, a penalty term is derived from the Gaussian approximation to compensate for this mismatch. This is a significant advance in the area of multilevel approximation where the likelihoods of symbol combinations with pseudosymbols are computed as if the pseudosymbols were truly in the constellation. The need for this penalty in the approximation is justified theoretically and its effectiveness is illustrated via computer simulations.

This paper is organized as follows. Section II describes the system model. The multilevel structure of the N-QAM constellation exploited by our algorithms is illustrated in Section III. The identification step of the two proposed multilevel SGA algorithms (MSGA-BFS, MSGA-DFS) are described on examples in Section IV and Section V respectively. In Section VI, simulation results are provided to illustrate the near-optimal performance of the proposed algorithms and we compare the complexities of the proposed algorithms with that of the SGA algorithm. In addition, a complexity reduction method via recursive update is explained in Appendix B.

II. System Model

Consider a spatial multiplexing MIMO system with N_T transmit antennas and $N_R \geq N_T$ receive antennas. At each time instant, N_T symbols $\mathbf{x} \stackrel{\text{def}}{=} [x_1, x_2, \ldots, x_{N_T}]^T$ ($[*]^T$ means transpose), taken from a modulation constellation $A = \{a_1, a_2, \ldots, a_N\}$, are transmitted from each antenna. Pertaining to them are N_R observations $\mathbf{y} \stackrel{\text{def}}{=} [y_1, y_2, \ldots, y_{N_R}]^T$. The relationship between $\mathbf{x}$ and $\mathbf{y}$ is :

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{H}$ is the $N_R \times N_T$ channel matrix with $h(i,j)$ as its (i,j)-th entry. The quantity $h(i,j)$ represents the channel gain from transmit antenna j to receive antenna i. The vector $\mathbf{n}$ is a $N_R \times 1$ vector of zero-mean complex circular symmetric Gaussian noise with covariance matrix $\sigma_n^2 \mathbf{I}$. We use $[*]^*$ and $[*]^H$ for the conjugate and transpose conjugate of a matrix or vector respectively.

The task of a space-time decoder is to estimate the transmitted symbol $\mathbf{x}$ from the observation $\mathbf{y}$ given the channel state information $\mathbf{H}$. More precisely, we are interested in the marginal posterior distributions $p(x_j|\mathbf{y}, \mathbf{H})$ for $j = 1, 2, \ldots, N_T$ (in what follows, conditioning on $\mathbf{H}$ will be implicit, and omitted).

The exact computation of the marginal posterior distributions $p(x_j|\mathbf{y})$ which requires an exhaustive search of all the possible symbol combinations can be efficiently approximated via the M most significant symbol combinations:

$$\begin{aligned} p(x_j|\mathbf{y}) &= \sum_{\mathbf{x}_{-j} \in D_{-j}} p(\mathbf{x}_{-j}, x_j|\mathbf{y}) \\ &\approx \sum_{m=1}^{M} p(x_1^{(m)}, \ldots, x_j, \ldots, x_{N_T}^{(m)}|\mathbf{y}), \end{aligned} \tag{2}$$

where $\mathbf{x}_{-j}$ refers to all the antennas except antenna j and D_{-j} is the set which contains the $N^{N_T - 1}$ possible values of $\mathbf{x}_{-j}$.

In the SGA algorithm [1], the identification of M most significant symbol combinations involves the computation and sorting of MN likelihoods for N_T steps.

In the following sections, we will explain the multilevel structure of the N-QAM constellation and develop two multilevel SGA algorithms with depth-first search and breadth-first search to identify M significant symbol combinations with this multilevel structure. The computation of the marginal symbol probabilities from those M identified symbol combinations is the same as Step 3 in [1, Section IV] and will be omitted here.

III. Multilevel Structure of the N-QAM Modulation Constellation

Fig. 1 describes the natural multilevel approximation of a 64-QAM symbol constellation A. The 64 dots represent the constellation A. We call this level $l = 1$. The 16 squares represent the 16-QAM approximation of constellation A used by our method at level $l = 2$. Effectively each square is the center of gravity of the four closest symbols from A (the dots). The four stars represent the 4-QAM approximation of the aforementioned 16-QAM constellation (the squares), and as a result the 4-QAM approximation of A at level $l = 3$. Note that these approximations define a quadtree, see Fig. 2. We will later on refer to parents and children on this tree.

More precisely, the definitions for the pseudosymbols are as follows. The set of pseudosymbols at the l-th level of approximation is defined as follows:

$$A_l \stackrel{\text{def}}{=} \{a_{l,1}, \ldots, a_{l,N_l}\}$$

for $l = 1, \ldots, L$ where $L \stackrel{\text{def}}{=} \log_4 N$ and $N_l \stackrel{\text{def}}{=} N4^{1-l}$. Note that at the lowest level where $l = 1$, A_1 is exactly A and $a_{1,k} = a_k, k = 1, \ldots N$. The pseudo symbol $a_{l,s}$ (here the parent) in set A_l is the mean value of 4 elements (here the

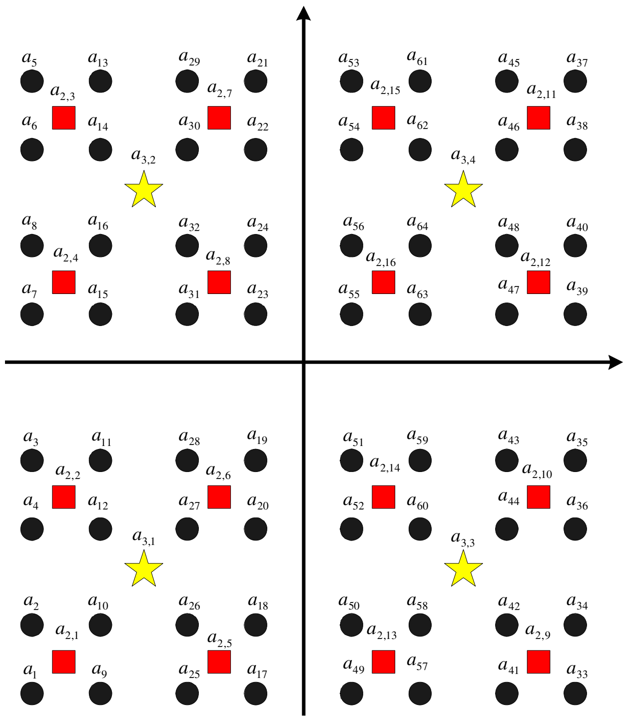


Fig. 1. Multilevel structure of 64QAM constellation.

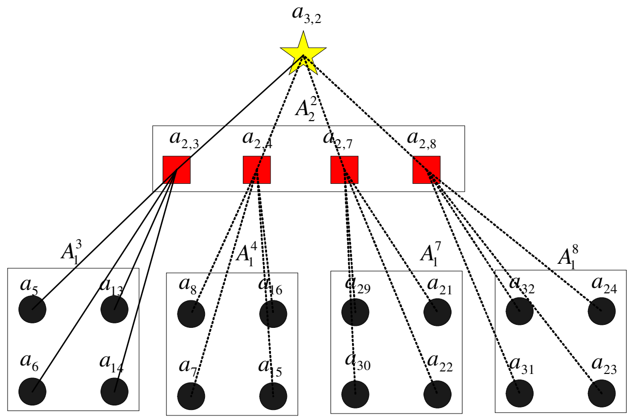


Fig. 2. An example of hierarchical structure of 64QAM constellation.

children) in a specific set A_{l-1}^s, which is a subset of the lower level set A_{l-1}:

$$a_{l,s} = 0.25 \sum_{a_{l-1,k} \in A_{l-1}^s} a_{l-1,k} \tag{3}$$

for $l = 2, \ldots, L, s = 1, \ldots, N_l$. Set A_l^s is a subset of A_l such that $\bigcup_k A_l^k = A_l$ and $A_l^k \bigcap A_l^n = \emptyset$ where $k, n = 1, \ldots, N_{l+1}$ and $k \neq n$.

As it is seen in the above definition of the pseudosymbols and the hierarchical structure of the constellation, a specific symbol $a_k \in A$ is only coupled to its ancestor $a_{l,s}$ at the l-th level [2]. To this end, we model the joint probability as:

$$p(a_k, a_{l,s}) \stackrel{\text{def}}{=} p(a_k)\mathbb{I}(a_k \texttt{ is a descendant of } a_{l,s}) \tag{4}$$

where $\mathbb{I}(.)$ is an indicator function for $k = 1, \ldots, N$, $l = 2, \ldots, L$ and $s = 1, \ldots, N_l$. Thus, the marginal probability for pseudo symbol $a_{l,s}$ is as follows:

$$p(a_{l,s}) = \sum_{a_k \in A} p(a_k)\mathbb{I}(a_k \texttt{ is a descendant of } a_{l,s}) \tag{5}$$

for $s = 1, \ldots, N_l$ and $l = 2, \ldots, L$.

IV. Multilevel SGA Detector With Depth-First Searching

A. Basic Idea

Suppose that we have symbol sequences up to the $(j-1)$-th antenna for system with 64QAM constellation. For the j-th antenna, the SGA algorithm evaluates $64 \times M$ symbol combinations for selection.

[2]The relationship can also be interpreted as follows: the conditional probability $p(a_{l,s}|a_k)$ is 1 if $a_{l,s}$ is an ancestor of a_k, or 0 otherwise.

In the MSGA-DFS algorithm, the selection consists of three steps. One first considers the approximation at level $l = L = 3$ (the stars), which results in $4 \times M$ likelihood estimations. The M significant symbol combinations are kept. Then we move on to the better approximation at level $l = 2$. We constrain the search of the significant symbols at level $l = 2$ to the children of the symbols selected at level $l = 3$, where only $4 \times M$ likelihoods are evaluated and sorted. Typically this is expected to reduce significantly the number of likelihood evaluations required. Then we repeat this down to the lowest approximation level $l = 1$. The total number of likelihoods that are evaluated in this process is $12 \times M$, which is only $3/16$ of that of the SGA algorithm.

A key element to the success of this process consists of taking into account the constellation approximation in the computation of the likelihood of symbol sequences with a pseudo symbol. We will describe the detailed algorithm and explain the effect of a multilevel constellation approximation theoretically in the next subsection.

B. Algorithm Description

Suppose that M significant combinations $\Theta_{j,l+1}^d \stackrel{\text{def}}{=} \{\left(x_1^{(m)}, \ldots, x_{j-1}^{(m)}, x_{j,l+1}^{(m)}\right), m = 1, 2, \ldots, M\}$ have been obtained for antenna $1, 2, \ldots, j$ at the $(l+1)$-th level. Then at the l-th level, only $4M$ pseudo symbol combinations $(x_1^{(m)}, \ldots, x_{j-1}^{(m)}, x_{j,l})$ are considered for $m = 1, \ldots, M$, where $x_{j,l} \in A_l^m$ is such that the mean of the elements in A_l^m is $x_{j,l+1}^{(m)}$, to form $\Theta_{j,1}^l$.

In order to select the pseudo symbol combinations, we must evaluate the approximate probabilities of $p(x_1^{(m)}, \ldots, x_{j-1}^{(m)}, x_{j,l}|\mathbf{y})$ for $m = 1, \ldots, M$ and $x_{j,l} \in A_l^m$. Define $\tilde{\mathbf{y}} \stackrel{\text{def}}{=} (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}$, then from Eq.(1), we have:

$$\tilde{\mathbf{y}} = \sum_{k=1}^{j-1} x_k\mathbf{e}_k + x_{j,l}\mathbf{e}_j + \underbrace{(x_j - x_{j,l})\mathbf{e}_j + \sum_{k=j+1}^{N_T} x_k\mathbf{e}_k + \tilde{\mathbf{n}}}_{\hat{\mathbf{n}}_{j,l}^d} \tag{6}$$

where the vector $\mathbf{e}_k$ is a column vector whose elements are all zeroes, but the k-th which is 1.

Before proceeding to the next step, we would like to compute the mean and variance of different items in the above equation. The Gaussian noise $\tilde{\mathbf{n}}$ has zero mean and variance $\mathbf{\Lambda} = \sigma_n^2(\mathbf{H}^H\mathbf{H})^{-1}$. The mean and variance of x_k w.r.t. a uniform distribution are zero and $\gamma = 1/N\sum_{a_s\in A}|a_s|^2$ respectively. The mean and variance of $x_j - x_{j,l}$ is zero and $\gamma_l = \gamma - \frac{1}{N_l}\sum_{a_{l,k}\in A_l}|a_{l,k}|^2$ respectively. The detailed computation of γ_l is given in Appendix A.

As a result, the mean and variance of $\hat{\mathbf{n}}_{j,l}^d$ (the pseudo-covariance vanishes in this case) are zero and $\mathbf{\Pi}_{j,l}^d$ where $\mathbf{\Pi}_{j,l}^d = \mathbf{\Pi}_j + \gamma_l \mathbf{e}_j\mathbf{e}_j^T$ and $\mathbf{\Pi}_j = \mathbf{\Lambda} + \gamma\sum_{k=j+1}^{N_T}\mathbf{e}_k\mathbf{e}_k^T$.

To evaluate the approximate probability of $p(x_1^{(m)},\ldots,x_{j-1}^{(m)},x_{j,l}|\mathbf{y})$, one models the distribution of $\hat{\mathbf{n}}_{j,l}^d$ as a moment-matched Gaussian distribution and uses the following approximation:

$$\begin{aligned}
& p(x_1^{(m)},\ldots,x_{j-1}^{(m)},x_{j,l}|\mathbf{y}) \\
\propto\ & p\left(\mathbf{y}|x_1^{(m)},\ldots,x_{j-1}^{(m)},x_{j,l}\right)p\left(x_{j,l}\right)\prod_{k=1}^{j-1}p(x_k^{(m)}) \\
\approx\ & \exp\Big(-\left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right)^H\left(\mathbf{\Pi}_{j,l}^d\right)^{-1} \\
& \left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right)\Big)p\left(x_{j,l}\right)\prod_{k=1}^{j-1}p(x_k^{(m)}) \\
\stackrel{\text{def}}{=}\ & \psi_m^d(x_{j,l})
\end{aligned} \tag{7}$$

where [3] $\mathbf{w}_{j-1}^{(m)} \stackrel{\text{def}}{=} \tilde{\mathbf{y}} - \sum_{k=1}^{j-1}x_k^{(m)}\mathbf{e}_k$.

It is worth commenting on that we use γ and γ_l in the above approximated likelihoods computation. The term γ accounts for the uncertainty introduced by the undetected symbols from antennas $j+1,\ldots,N_T$, which has been suggested in the SGA algorithm. The term γ_l accounts for the additional uncertainty introduced by using the pseudosymbols $x_{j,l}$, which has never been proposed before (even in the multilevel literature [13] [14] [15] [16]).

Taking γ_l into consideration for computing the likelihoods is an important methodological contribution to the multilevel literature. In order to further explain the effect of this term from a theoretical perspective, we can obtain the following equation using the Sherman-Morrison-Woodbury formula:

$$\begin{aligned}
& \left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right)^H(\mathbf{\Pi}_{j,l}^d)^{-1}\left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right) \\
=\ & \left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right)^H\left(\mathbf{\Pi}_{j,1}^d + \gamma_l\mathbf{e}_j\mathbf{e}_j^T\right)^{-1} \\
& \left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right) \\
=\ & \left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right)^H(\mathbf{\Pi}_{j,1}^d)^{-1}\left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right) \\
& -\varpi_{j,l}^{(m)},
\end{aligned} \tag{8}$$

$$\begin{aligned}
\varpi_{j,l}^{(m)} \stackrel{\text{def}}{=}\ & \left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right)^H\left(\gamma_l^{-1} + [(\mathbf{\Pi}_{j,1}^d)^{-1}]_{(j,j)}\right)^{-1} \\
& \left([(\mathbf{\Pi}_{j,1}^d)^{-1}]_{(:,j)}\left([(\mathbf{\Pi}_{j,1}^d)^{-1}]_{(:,j)}\right)^H\right) \\
& \left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right)
\end{aligned}$$

where $[\left(\mathbf{\Pi}_{j,1}^d\right)^{-1}]_{(:,j)}$ and $[\left(\mathbf{\Pi}_{j,1}^d\right)^{-1}]_{(j,j)}$ denote the j-th column and diagonal element of matrix $\left(\mathbf{\Pi}_{j,1}^d\right)^{-1}$ respectively.

For $l = 1$ where $\gamma_1 = 0$, $x_{j,1} = x_1 \in A$, $\mathbf{\Pi}_{j,1}^d = \mathbf{\Pi}_j$ and $\varpi_{j,l}^{(m)} = 0$, the above equation is identical to that corresponding to the SGA algorithm. If $l \neq 1$ and hence $\gamma_l \neq 0$, the extra penalty term $\varpi_{j,l}^{(m)}$, which is derived from the Gaussian approximation, compensates for the mismatch between the multilevel pseudosymbols $x_{j,l} \in A_l$ and the actual transmitted symbols which take values in A. The beneficial effect of taking this penalty term into consideration is confirmed in simulations.

Finally the M symbol combinations with the largest $\psi_m^d(x_{j,l})$ are selected among the $4M$ possible symbol combinations, resulting in a new set $\Theta_{j,l}^d$. For the last step $l = 1$ for which $\gamma_1 = 0$ and $p\left(x_{j,1}\right) = p(x_j)$, a set of M significant symbol combinations $\Theta_{j,1}^d = \{\left(x_1^{(m)},\ldots,x_j^{(m)}\right), m = 1,\ldots,M\}$ is obtained. An example is given in the next subsection to illustrate the MSGA-DFS identification step.

C. Summary of the MSGA-DFS Identification Step

1) Compute the zero forcing output $\tilde{\mathbf{y}}$ and initialize the set of symbol combinations $\Theta_{0,1}^d = \emptyset$, $\tilde{M} = 0$. Compute $\psi_0^d(x_{1,1})$ for $x_{1,1} \in A$ and select the $\tilde{M} = \min(M,N)$ largest ones [4] for set $\Theta_{1,1}^d$. For $1 < j \leq N_T$,
   a) Compute $\psi_m^d(x_{j,L})$ according to Eq. (**??**) and Eq. (8) for all the elements in $\Theta_{j-1,1}^d = \{\left(x_1^{(m)},\ldots,x_{j-1}^{(m)}\right), m = 1,\ldots,\tilde{M}\}$ and $x_{j,L} \in A_L$,
   b) Select $\tilde{M} = \min(M, 4^{(j-1)L+1})$ symbol combinations which have the largest $\psi_m^d(x_{j,L})$ for set $\Theta_{j,L}^d = \{\left(x_1^{(m)},\ldots,x_{j-1}^{(m)},x_{j,L}^{(m)}\right), m = 1,\ldots,\tilde{M}\}$. For $l = L-1,\ldots,1$,
      i) Compute $\psi_m^d(x_{j,l})$ according to Eq. (**??**) for all the elements in $\Theta_{j,l+1}^d$ and $x_{j,l} \in A_l^m$ (the elements in set A_l^m are children of $x_{j,l+1}^{(m)}$) for $m = 1,\ldots,\tilde{M}$,
      ii) Select the $\tilde{M} = \min(M, 4^{jL-l+1})$ symbol combinations which have the largest $\psi_m^d(x_{j,l})$ for $\Theta_{j,l}^d = \{\left(x_1^{(m)},\ldots,x_{j-1}^{(m)},x_{j,l}^{(m)}\right), m = 1,\ldots,\tilde{M}\}$.

V. Multilevel SGA Detector With Breadth-First Searching

A. Basic Idea

The multilevel algorithms proposed in the last section (MSGA-DFS) as well as in the literature (the multilevel mixture Kalman filter [16] and the ITS detector [13] [14] [15]) are all based on the depth-first search described in the last section.

In this section, we propose a breadth-first search based multilevel SGA algorithm (MSGA-BFS). The MSGA-BFS algorithm consists of considering the multilevel approximation

[3] For notational simplicity, we drop j from $\psi_m^d(.)$. The notation $\psi_0^d(.)$ is used for the first antenna where no previous symbol combinations are available.

[4] For the first antenna, we compute the approximate likelihoods of all N possible symbols without the DFS based multilevel approximation.

for the pseudosymbols from all the antennas at level $l = L$ first, and then refine the approximations at levels $l = L-1, \ldots, 1$ sequentially.

Again we use a 64QAM constellation for the purpose of illustration. Firstly, we consider all the possible pseudosymbol combinations at level 3. There are 4^{N_T} pseudo symbol combinations (4 possible pseudosymbols per antenna) in total. The sequential identification procedure proposed in the SGA algorithm is employed to identify the set of M significant pseudo symbol combinations taking into account the multi-level approximation, i.e. the term γ_L as explained in the last section. The total cost for this identification procedure comes from the computing and sorting of $4M$ likelihoods for N_T steps. Then we move down to the next level with the set of the 3rd level pseudo symbol combinations defined as follows:

$$\Theta^b_{N_T,L} \stackrel{\text{def}}{=} \{\left(x^{(m)}_{1,L}, \ldots, x^{(m)}_{N_T,L}\right), m = 1, \ldots, M\}$$

At the second level where $l = 2$, we evaluate the $4M$ symbol combinations for antennas $j = 1, \ldots, N_T$ sequentially as follows. For $j = 1$, we evaluate likelihoods of the $4M$ symbol combinations $(x_{1,l}, x^{(m)}_{2,l+1}, \ldots, x^{(m)}_{N_T,l+1})$ with the constraint that $x_{1,l}$ is a child of $x^{(m)}_{1,l+1}$ for $m = 1, \ldots, M$, and only keep the M pseudo symbol combinations with the largest likelihoods for the next step where $j = 2$. The evaluation of the likelihoods here also takes into consideration the mismatch between the pseudosymbols and true symbols (the term γ_l). The above evaluation and selection step is repeated until the last antenna is reached where $j = N_T$.

The above procedure is repeated for $l = 1$ and finally the M significant symbol combinations can be obtained for the computation of the posterior symbol probabilities. The total number of likelihoods that are computed and sorted is the same as that of the MSGA-DFS algorithm.

In the next section, a detailed description of the MSGA-BFS algorithm is presented.

B. Algorithm Description

1) Selection Procedure for the Highest Level ($l = L$): The aim of the identification procedure at the highest level $l = L$ is to select the set of M pseudo symbol combinations:

$$\Theta^b_{j,L} = \{\left(x^{(m)}_{1,L}, \ldots, x^{(m)}_{j,L}\right), m = 1, \ldots, M\}$$

from all the possible 4^{N_T} pseudo symbol combinations for $j = 1, \ldots, N_T$. The identification procedure is similar to that of the SGA algorithm where in the j-th step, the likelihoods of $4M$ symbol combinations are computed and sorted to select the M highly probable ones for the next step, given the set of pseudo symbol combinations from the $j-1$-th step defined as follows:

$$\Theta^b_{j-1,L} = \{\left(x^{(m)}_{1,L}, \ldots, x^{(m)}_{j-1,L}\right), m = 1, \ldots, M\}.$$

To compute the required likelihoods, we first rewrite the system model as follows:

$$\tilde{\mathbf{y}} = \sum_{k=1}^{j} x_{k,L}\mathbf{e}_k + \underbrace{\sum_{k=1}^{j}(x_k - x_{k,L}) + \sum_{k=j+1}^{N_T} x_k\mathbf{e}_k + \tilde{\mathbf{n}}}_{\hat{\mathbf{n}}^b_{j,L}}. \quad (9)$$

where the mean and variance of x_k w.r.t. the uniform distribution are zero and γ respectively. The mean and variance of $x_k - x_{k,L}$ are zero and γ_L respectively. Thus, the variance [5] of $\hat{\mathbf{n}}^b_{j,L}$ is $\mathbf{\Pi}^b_{j,L} \stackrel{\text{def}}{=} \mathbf{\Lambda} + \gamma_L \sum_{k=1}^{j} \mathbf{e}_k\mathbf{e}_k^T + \gamma \sum_{k=j+1}^{N_T} \mathbf{e}_k\mathbf{e}_k^T$.

Now one models the distribution of $\hat{\mathbf{n}}^b_{j,L}$ as a moment-matched Gaussian distribution and calculates $p\left(x^{(m)}_{1,L}, \ldots, x^{(m)}_{j-1,L}, x_{j,L}|\mathbf{y}\right)$ as follows:

$$\begin{aligned}
& p\left(x^{(m)}_{1,L}, \ldots, x^{(m)}_{j-1,L}, x_{j,L}|\mathbf{y}\right) \\
\propto\ & p\left(\mathbf{y}|x^{(m)}_{1,L}, \ldots, x^{(m)}_{j-1,L}, x_{j,L}\right) p(x_{j,L}) \prod_{k=1}^{j-1} p\left(x^{(m)}_{k,L}\right) \\
\approx\ & \exp\Big(-\left(\mathbf{w}^{(m)}_{j-1,L} - x_{j,L}\mathbf{e}_j\right)^H \left(\mathbf{\Pi}^b_{j,L}\right)^{-1} \\
& \left(\mathbf{w}^{(m)}_{j-1,L} - x_{j,L}\mathbf{e}_j\right)\Big) p(x_{j,L}) \prod_{k=1}^{j-1} p\left(x^{(m)}_{k,L}\right) \\
\stackrel{\text{def}}{=}\ & \psi^b_m(x_{j,L}) \qquad (10)
\end{aligned}$$

where $\mathbf{w}^{(m)}_{j-1,L} \stackrel{\text{def}}{=} \tilde{\mathbf{y}} - \sum_{k=1}^{j-1} x^{(m)}_{k,L}\mathbf{e}_k$.

Then M symbol combinations with largest $\psi^b_m(x_{j,L})$ are selected for $\Theta^b_{j,L}$. This procedure is repeated for $j = 1, \ldots, N_T$ until $\Theta^b_{N_T,L}$ is formed.

2) Selection Procedure for the l-th Level: At the l-th level for $l = L-1, \ldots, 2, 1$, the aim is to identify the M significant pseudo symbol combinations $\Theta^b_{N_T,l} = \{\left(x^{(m)}_{1,l}, \ldots, x^{(m)}_{N_T,l}\right), m = 1, \ldots, M\}$ using the pseudo symbol combinations $\Theta^b_{N_T,l+1} = \{\left(x^{(m)}_{1,l+1}, \ldots, x^{(m)}_{N_T,l+1}\right), m = 1, \ldots, M\}$ identified in the last step.

This identification step can be decomposed into $j = 1, \ldots, N_T$ steps where in the j-th step, $4M$ approximate posterior probabilities $p\left(x^{(m)}_{1,l}, \ldots, x^{(m)}_{j-1,l}, x_{j,l}, x^{(m)}_{j+1,l+1}, \ldots, x^{(m)}_{N_T,l+1}|\mathbf{y}\right)$ are computed with the constraints that $x_{j,l} \in A^{(m)}_l$ for $m = 1, \ldots, M$. Then M pseudo symbol combinations with largest approximate probabilities are selected for the $j+1$-th step and the set $\Theta^b_{j,l} = \{\left(x^{(m)}_{1,l}, \ldots, x^{(m)}_{j-1,l}, x^{(m)}_{j,l}, x^{(m)}_{j+1,l+1}, \ldots, x^{(m)}_{N_T,l+1}\right), m = 1, \ldots, M\}$ is obtained.

The approximated posterior probabilities $p\left(x^{(m)}_{1,l}, \ldots, x^{(m)}_{j-1,l}, x_{j,l}, x^{(m)}_{j+1,l+1}, \ldots, x^{(m)}_{N_T,l+1}|\mathbf{y}\right)$ can be computed via a Gaussian approximation. First we rewrite the decorrelating model as follows:

$$\begin{aligned}
\tilde{\mathbf{y}} =\ & \sum_{k=1}^{j-1} x_{k,l}\mathbf{e}_k + x_{j,l}\mathbf{e}_j + \sum_{k=j+1}^{N_T} x_{k,l+1}\mathbf{e}_k \\
& \underbrace{+ \sum_{k=1}^{j}(x_k - x_{k,l})\mathbf{e}_k + \sum_{k=j+1}^{N_T}(x_k - x_{k,l+1})\mathbf{e}_k + \tilde{\mathbf{n}}}_{\hat{\mathbf{n}}^b_{j,l}}. \qquad (11)
\end{aligned}$$

[5]The terms $\mathbf{\Lambda}$, γ and γ_L are defined in the previous section.

where both $x_k - x_{k,l}$ and $x_k - x_{k,l+1}$ have zero means and their variance are γ_l and γ_{l+1} respectively. Thus the variance of $\hat{\mathbf{n}}^b_{j,l}$ is $\mathbf{\Pi}^b_{j,l} \stackrel{\text{def}}{=} \mathbf{\Lambda} + \gamma_l \sum_{k=1}^{j} \mathbf{e}_k\mathbf{e}_k^T + \gamma_{l+1}\sum_{k=j+1}^{N_T} \mathbf{e}_k\mathbf{e}_k^T$.

Now one models the distribution of $\hat{\mathbf{n}}^b_{j,l}$ as a moment-matched Gaussian distribution:

$$\begin{aligned} & p\left(x^{(m)}_{1,l},\ldots,x^{(m)}_{j-1,l},x_{j,l},x^{(m)}_{j+1,l+1},\ldots,x^{(m)}_{N_T,l+1}|\mathbf{y}\right) \\ \propto \quad & p\left(\mathbf{y}|x^{(m)}_{1,l},\ldots,x^{(m)}_{j-1,l},x_{j,l},x^{(m)}_{j+1,l+1},\ldots,x^{(m)}_{N_T,l+1}\right) \\ & p(x_{j,l})\prod_{k=1}^{j-1} p\left(x^{(m)}_{k,l}\right)\prod_{k=j+1}^{N_T} p\left(x^{(m)}_{k,l+1}\right) \\ \approx \quad & \exp\Big(-\left(\mathbf{w}^{(m)}_{j-1,l} - \left(x_{j,l}-x^{(m)}_{j,l+1}\right)\mathbf{e}_j\right)^H \left(\mathbf{\Pi}^b_{j,l}\right)^{-1} \\ & \left(\mathbf{w}^{(m)}_{j-1,l} - \left(x_{j,l}-x^{(m)}_{j,l+1}\right)\mathbf{e}_j\right)\Big) \\ & p(x_{j,l})\prod_{k=1}^{j-1} p\left(x^{(m)}_{k,l}\right)\prod_{k=j+1}^{N_T} p\left(x^{(m)}_{k,l+1}\right) \\ \stackrel{\text{def}}{=} \quad & \psi^b_m(x_{j,l}) \qquad (12) \end{aligned}$$

where $\mathbf{w}^{(m)}_{j-1,l} = \tilde{\mathbf{y}} - \sum_{k=1}^{j-1} x^{(m)}_{k,l}\mathbf{e}_k - \sum_{k=j}^{N_T} x^{(m)}_{k,l+1}\mathbf{e}_k$.

The above procedure is repeated for $l = L-1,\ldots,2,1$ and finally $\Theta^b_{N_T,1}$ can be obtained for the computation of the marginal symbol probabilities.

C. Summary of the MSGA-BFS Identification Step

1) Compute the zero forcing output $\tilde{\mathbf{y}}$ and initialize the set of symbol combinations $\Theta^b_{0,L} = \emptyset$, $\tilde{M} = 0$. For $j = 1$ compute $\psi^b_0(x_{1,L})$ for $x_{1,L} \in A_L$.
   a) For $1 < j \le N_T$, compute $\psi^b_m(x_{j,L})$ for all the elements in $\Theta^b_{j-1,L} = \{\left(x^{(m)}_{1,L},\ldots,x^{(m)}_{j-1,L}\right), m = 1,\ldots,\tilde{M}\}$ and $x_{j,L} \in A_L$ according to Eq. (10).
   b) Select the $\tilde{M} = \min(M, 4^j)$ symbol combinations which have the largest $\psi^b_m(x_{j,L})$ for $\Theta^b_{j,L} = \{\left(x^{(m)}_{1,L},\ldots,x^{(m)}_{j,L}\right), m = 1,\ldots,\tilde{M}\}$.
2) For $l = L-1,\ldots,2,1$, Set $\Theta^b_{0,l} = \Theta^b_{N_T,l+1}$:
   a) For $1 \le j \le N_T$, compute $\psi^b_m(x_{j,l})$ according to Eq. (12) for $4\tilde{M}$ symbol combinations $\left(x^{(m)}_{1,l},\ldots,x^{(m)}_{j-1,l},x_{j,l},x^{(m)}_{j+1,l+1},\ldots,x^{(m)}_{N_T,l+1}\right)$ where $x_{j,l} \in A^m_l$ (the elements in set A^m_l are children of $x^{(m)}_{j,l+1}$) and
   $$\left(x^{(m)}_{1,l},\ldots,x^{(m)}_{j-1,l},x^{(m)}_{j,l+1},x^{(m)}_{j+1,l+1},\ldots,x^{(m)}_{N_T,l+1}\right)$$
   $\in \Theta^b_{j-1,l}$ for $m = 1,\ldots,\tilde{M}$,
   b) Select the $\tilde{M} = \min(M, 4^{(L-l)N_T+j})$ symbol combinations which have the largest $\psi^b_m(x_{j,l})$ for $\Theta^b_{j,l} = \{\left(x^{(m)}_{1,l},\ldots,x^{(m)}_{j,l},x^{(m)}_{j+1,l+1},\ldots,x^{(m)}_{N_T,l+1}\right), m = 1,\ldots,\tilde{M}\}$.

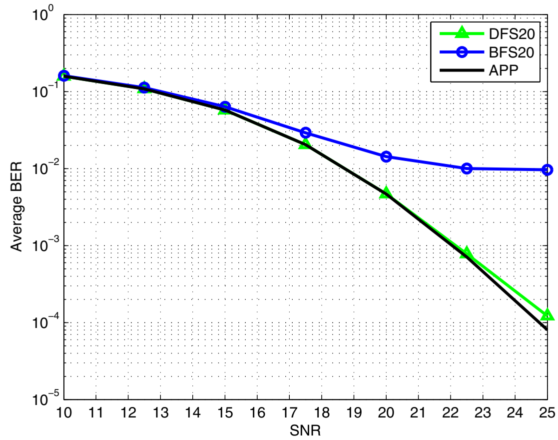


Fig. 3. Uncoded BER performance of the DFS and BFS algorithms for a 16QAM, 4×4 serial system. Both of the algorithms fail to take into account the additional uncertainty from the use of pseudosymbols.

VI. Simulation Results

In this section, we demonstrate the near-optimal performance of the proposed MSGA-DFS and MSGA-BFS algorithms in various scenarios and the importance of the correction term of the multilevel approximation. In all our simulations, we set $N_T = N_R = 4$ and consider 16QAM/64QAM modulation systems with 1152 bits per frame before channel coding. The SNR is defined as $E\{||\mathbf{Hx}||^2\}/E\{||\mathbf{n}||^2\} = \gamma N_T/\sigma_n^2$.

A $1/2$ rate turbo code with polynomials 7 and 5 is used at the transmitter and a BCJR channel decoder with 4 iterations is used at the receiver. There are no outer iterations, i.e. the MIMO decoder processes the data only once. For each SNR we randomly generate 5×10^4 channel realizations, which were processed by all algorithms.

A. Effect of the Gaussian Approximation

The effect of the Gaussian approximation is investigated via comparison with two algorithms termed the depth-first search (DFS) and breadth-first search (BFS) algorithms. Both algorithms (DFS and BFS) are with Gaussian approximation variance term γ, but without proper Gaussian approximation for multilevel pseudo symbols.

1) The DFS algorithm is similar to the MSGA-DFS algorithm described in section IV-C except that the variance terms and $\gamma_l, l = L,\ldots,2$ are set to 0.
2) The BFS algorithm is similar to the MSGA-BFS algorithm described in section V-C except that the variance terms $\gamma_l, l = L,\ldots,2$ are set to 0.

Fig. 3 and Fig. 4 shows the uncoded BER performance of DFS and BFS with $M = 20$ for a 16QAM, 4×4 system and $M = 40$ for a 64QAM, 4×4 system respectively. It is seen that both algorithms (DFS,BFS) experience error floors and perform worse than that of SD in high SNR region (the interference is significant) and for 64QAM constellation (the

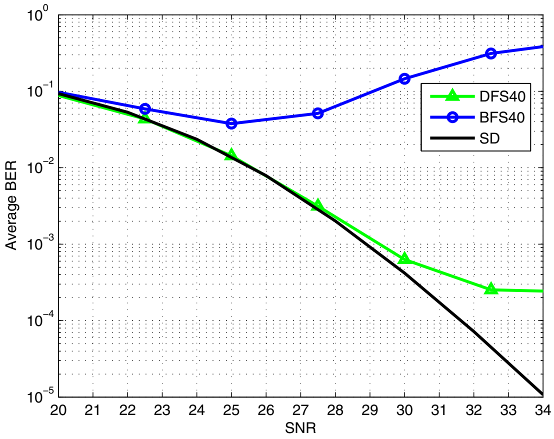


Fig. 4. Uncoded BER performance of the DFS and BFS algorithms for a 64QAM, 4×4 serial system. Both of the algorithms fail to take into account the additional uncertainty from the use of pseudosymbols.

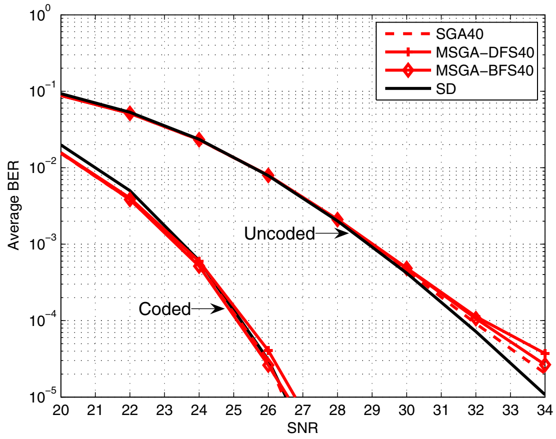


Fig. 6. BER performance of the MSGA-DFS and MSGA-BFS algorithms for a 64QAM, 4×4 system.

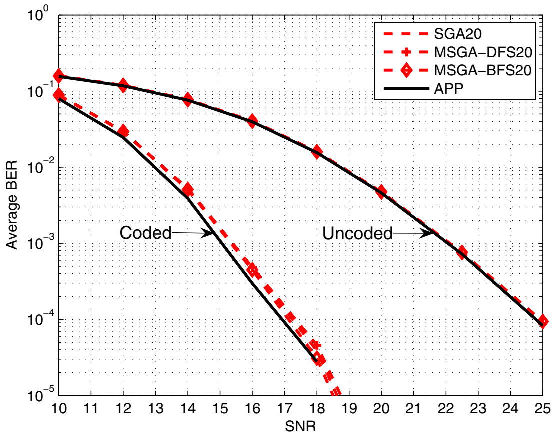


Fig. 5. BER performance of the MSGA-DFS and MSGA-BFS algorithms for a 16QAM, 4×4 system.

variance term γ_l is large) [6].

The variance term γ_l, $\left(\mathbf{\Pi}^d_{j,l}\right)^{-1}$ and $\left(\mathbf{\Pi}^b_{j,l}\right)^{-1}$ can be computed once for a block with constant $\mathbf{H}$ over the block. The amount of additional complexity required by considering variance term γ_l is 1 addition for each computation of $\psi^d_m(x_{j,l})$ in Eq. (14) for the MSGA-DFS algorithm and $\psi^b_m(x_{j,l})$ in Eq. (16) and Eq. (17) for the MSGA-DFS algorithm. As a result, the additional complexity is negligible (less than 1%) in total.

B. Performance Comparison of MSGA-DFS and MSGA-BFS Algorithms

The uncoded and coded performance of the a posterior probability (APP), complex formulation PDA algorithm (CPDA) [10], SGA, MSGA-DFS and MSGA-BFS algorithms with $M = 20$ is presented in Fig. 5 for a 16QAM 4×4 system. It is seen that the performance of MSGA-DFS and MSGA-BFS approaches that of SGA in both the uncoded and coded cases with $M = 20$ respectively. Both MSGA-DFS and MSGA-BFS with $M = 20$ works nearly as well as the APP.

Fig. 6 shows performance of the Max-log SD [12][7], SGA, MSGA-DFS and MSGA-BFS for 64QAM, 4×4 system. In the uncoded case, performance of the MSGA and SGA algorithms is slightly worse than that of SD for high SNR levels. It is also noticed that performance of the MSGA-BFS algorithms is slightly better than that of the MSGA-DFS algorithm (with the same M). In the coded case, the performance of the SD algorithm is similar to that of the SGA and MSGA algorithms with $M = 40$.

C. Complexity Comparison

Table I shows the algebraic complexity of the SGA algorithm and MSGA algorithms (MSGA-DFS, MSGA-BFS)[8] for selection of M symbol combinations for one antenna. It is seen that the comparisons required in the MSGA algorithms should be significantly smaller than that in SGA algorithm. Table II shows the number of real operations (MUL+ADD+COMP) per time instant for the SGA algorithm, the SD algorithm [12] and the proposed MSGA-DFS/MSGA-BFS algorithms for 4×4, 16QAM/64QAM systems. The number of operations of the SD algorithm is averaged over 1000 channel realizations with SNR=16 dB for 16QAM constellation and SNR=24dB for 64QAM constellation respectively.

It is noticed that the complexities of the three SGA based algorithms (SGA,MSGA-DFS,MSGA-BFS) are much lower than the average complexity of the SD algorithm. The complexity of the MSGA algorithms is only around 81% and 48% of that of the original SGA algorithm for the 16QAM and 64QAM systems respectively.

[6]It is observed in the simulations that the likelihoods of symbol combinations becomes smaller and smaller (near to numerical precision in Matlab) in high SNR levels for the BFS algorithm with 64QAM . Thus, the selection step becomes inaccurate and the performance of the BFS algorithm is severely degraded with increasing SNR.

[7]The SD algorithm [12] used here is a benchmark which has been shown superior to standard list SD [4]. There are many further improvements about SD with pre-processing and post-processing methods proposed recently. But we opt for a standard implementation.

[8]The recursive updating method proposed in Appendix B is used to reduce the complexity of the SGA, MSGA-DFS and MSGA-BFS algorithms. It is assumed that the heap sorting algorithm is used for partial sorting in the SGA, SGA-DFS and SGA-BFS algorithms, which has a average complexity of $\mathcal{O}\left(MN\log(MN)\right)$

TABLE I

OPERATIONS COMPARISON PER IDENTIFICATION PROCEDURE FOR ONE ANTENNA

	ADD	MUL	COMP(average)
SGA	$M(4N_T+2N+2)$	$M(4N_T+3N+3)$	$\mathcal{O}(MN\log(MN))$
MSGA-DFS,MSGA-BFS	$ML(4N_T+10)$	$ML(4N_T+15)$	$\mathcal{O}(L4M\log(4M))$

TABLE II

OPERATIONS COMPARISON PER TIME INSTANT FOR A 4×4,16QAM/64QAM MIMO SYSTEM

	16QAM				64QAM			
	SGA20	MSGA-DFS20	MSGA-BFS20	SD(16dB)	SGA40	MSGA-DFS40	MSGA-BFS40	SD(24dB)
ADD	6930	6050	6076	198088	34946	26786	26862	584778
MUL	8887	8587	8618	148566	53411	38251	38349	438584
COMP	3878	1143	1143	3096	57616	5654	5654	9137
Total	19395	15780	15837	349750	145973	70691	70865	1032500

VII. CONCLUSIONS

In this paper, two multilevel SGA algorithms with depth-first searching (MSGA-DFS) and breadth-first searching (MSGA-BFS) are proposed to reduce the complexity of the original SGA algorithm for near-optimal detection of MIMO system with higher order QAM constellations (16QAM/64QAM).

The two methods exploit the multilevel structure of QAM constellations to reduce the effect of large constellation size on computation and sorting. Simulation results demonstrate that both of the algorithms can achieve near-optimal (APP) performance in both coded and uncoded systems for a complexity which is only around 81% and 48% of that of the original SGA algorithm for MIMO system with 4 transmit and receive antennas and 16QAM, 64QAM modulation constellations, respectively.

VIII. APPENDIX

A. Computation of γ_l

The detailed computation of γ_l is given as follows:

$$\begin{aligned}
\gamma_l &= Var(x_j - x_{j,l}) \\
&= E|x_j|^2 + E|x_{j,l}|^2 - E(x_j x_{j,l}^*) - E(x_j^* x_{j,l}) \\
&= \gamma + \frac{1}{N_l}\sum_{a_{l,k}\in A_l}|a_{l,k}|^2 \\
&- \sum_{x_{j,l}\in A_l} x_{j,l}^* \sum_{x_j\in A} x_j p(x_j, x_{j,l}) \\
&- \sum_{x_{j,l}\in A_l} x_{j,l} \sum_{x_j\in A} x_j^* p(x_j, x_{j,l}) \\
&= \gamma + \frac{1}{N_l}\sum_{a_{l,k}\in A_l}|a_{l,k}|^2 - 2*\frac{4^{l-1}}{N}\sum_{a_{l,k}\in A_l}|a_{l,k}|^2 \\
&= \gamma - \frac{1}{N_l}\sum_{a_{l,k}\in A_l}|a_{l,k}|^2.
\end{aligned}$$

where the joint probability $p(x_j, x_{l,k})$ is given in Eq. (4), the mean and variance of $x_{j,l}$ w.r.t. a uniform distribution are zero and $\frac{1}{N_l}\sum_{a_{l,k}\in A_l}|a_{l,k}|^2$ respectively for $l = 1,\ldots,L$. These calculations can straightforwardly be altered in order to consider the case where a non-uniform prior is used, such as in a Turbo decoding framework.

B. Complexity Reductions

In this section, we propose recursively update methods to reduce the complexity of the proposed MSGA-DFS and MSGA-BFS algorithms via the matrix inversion lemma.

1) Recursive Update of $\psi_m^d(x_{j,l})$: The initialization of the MSGA-DFS algorithm for $j=1$ is the same as for the SGA algorithm which computes $\psi_0^d(x_{1,1})$ and is as follows:

$$\begin{aligned}
\psi_0^d(x_{1,1}) &= \exp\left(\begin{array}{c} -(\tilde{\mathbf{y}} - x_{1,1}\mathbf{e}_1)^H \left(\mathbf{\Pi}_{1,1}^d\right)^{-1} \\ (\tilde{\mathbf{y}} - x_{1,1}\mathbf{e}_1) \end{array}\right) \\
&\quad p(x_{1,1}) \\
&\propto \exp\Big(2\Re\left(x_{1,1}\tilde{\mathbf{y}}^H[\left(\mathbf{\Pi}_{1,1}^d\right)^{-1}]_{(:,1)}\right) \\
&\quad - |x_{1,1}|^2 [\left(\mathbf{\Pi}_{1,1}^d\right)^{-1}]_{(1,1)}\Big)p(x_{1,1}) \qquad (13)
\end{aligned}$$

with $x_{1,1} = x_1 \in A$ and $\mathbf{\Pi}_{1,1}^d = \mathbf{\Pi}_1$.

Then $\tilde{M} = \min(M,N)$ symbols $x_{1,1}^{(m)}$ with the largest $\psi_0^d(x_{1,1})$ for $m = 1,\ldots,\tilde{M}$ are selected and stored for next step.

With $\Theta_{j-1,1}^d = \{(x_1^{(m)},\ldots,x_{j-1}^{(m)})\}$ and

$$\psi_m^d\left(x_{j-1,1}^{(m)}\right) = \exp\left(-\left(\mathbf{w}_{j-1}^{(m)}\right)^H \left(\mathbf{\Pi}_{j-1,1}^d\right)^{-1}\mathbf{w}_{j-1}^{(m)}\right)\prod_{k=1}^{j-1} p\left(x_k^{(m)}\right),$$

obtained from the last step, the $\psi_m^d(x_{j,l})$ in Eq. (13) can be computed recursively for $l = L,\ldots,1, j = 2,\ldots,N_T$ as follows:

$$\begin{aligned}
&\psi_m^d(x_{j,l}) \\
&= \exp\Big(-\left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right)^H \left(\mathbf{\Pi}_{j,l}^d\right)^{-1} \\
&\quad \left(\mathbf{w}_{j-1}^{(m)} - x_{j,l}\mathbf{e}_j\right)\Big)p(x_{j,l})\prod_{k=1}^{j-1}p(x_k^{(m)}) \\
&= \psi_m^d\left(x_{j-1,1}^{(m)}\right)\exp\Big(-\zeta_{j,l}^d\left|\eta_{j,l}^{(m)}\right|^2 + 2\Re\left(x_{j,l}\eta_{j,l}^{(m)}\right) \\
&\quad -|x_{j,l}|^2[\left(\mathbf{\Pi}_{j,l}^d\right)^{-1}]_{(j,j)}\Big)p(x_{j,l}), \qquad (14)
\end{aligned}$$

$$\eta_{j,l}^{(m)} = \left(\mathbf{w}_{j-1}^{(m)}\right)^H [(\mathbf{\Pi}_{j,l}^d)^{-1}]_{(:,j)},$$

$$\zeta_{j,l}^d = \left((\gamma_{l+1} - \gamma_l)^{-1} + [(\mathbf{\Pi}_{j,l}^d)^{-1}]_{(j,j)}\right)^{-1}.$$

The following matrix inversion lemma is used in the above equation:

$$\left(\mathbf{\Pi}_{j,l}^d\right)^{-1} = \left(\mathbf{\Pi}_{j-1,1}^d\right)^{-1} + \zeta_{j,l}^d [(\mathbf{\Pi}_{j,l}^d)^{-1}]_{(:,j)} \left([(\mathbf{\Pi}_{j,l}^d)^{-1}]_{(:,j)}\right)^H$$

for $j = 2, \ldots, N_T$.

Note that $\psi_m^d\left(x_{j-1,1}^{(m)}\right)$ is stored in the last level $l = 1$ for the $j-1$-th antenna and the computation of $\eta_j^{(m)}$ dominates the complexity. So the total complexity of the MSGA-DFS identification step excluding partial sorting and block operations (for constant channels $\mathbf{H}$ over one block) is $\mathcal{O}(MN_T^2)$.

2) Recursive Update of $\psi_m^b(x_{j,l})$: The initialization of the MSGA-BFS identification procedure is different to that of the MSGA-DFS algorithm. The computation of $\psi_m^b(x_{1,L})$ for $x_{j,L} \in A_L$ is as follows:

$$\begin{aligned}
&\psi_0^b(x_{1,L}) \\
&\quad= \exp\left(-(\tilde{\mathbf{y}} - x_{1,L}\mathbf{e}_1)^H \left(\mathbf{\Pi}_{1,L}^b\right)^{-1} (\tilde{\mathbf{y}} - x_{1,L}\mathbf{e}_1)\right) \\
&\qquad p(x_{1,L}) \\
&\quad\propto \exp\left(2\Re\left(x_{1,L}\tilde{\mathbf{y}}^H [\left(\mathbf{\Pi}_{1,L}^b\right)^{-1}]_{(:,1)}\right)\right. \\
&\qquad \left. - |x_{1,L}|^2 [\left(\mathbf{\Pi}_{1,L}^b\right)^{-1}]_{(1,1)}\right) p(x_{1,L}) \qquad (15)
\end{aligned}$$

with $\mathbf{\Pi}_{1,L}^d = \mathbf{\Pi}_1 + \gamma_L \mathbf{e}_1 \mathbf{e}_1^T$. Then $\tilde{M} = \min(M, 4)$ symbols $x_{1,L}^{(m)}$ with largest $\psi_0^b(x_{1,L})$ for $m = 1, \ldots, \tilde{M}$ are selected and stored for next step.

With $\Theta_{j-1,L}^b = \{(x_{1,L}^{(m)}, \ldots, x_{j-1,L}^{(m)})$ and

$$\begin{aligned}
\psi_m^b\left(x_{j-1,L}^{(m)}\right) = &\prod_{k=1}^{j-1} p\left(x_{k,L}^{(m)}\right) \\
&\exp\left(-\left(\mathbf{w}_{j-1,L}^{(m)}\right)^H \left(\mathbf{\Pi}_{j-1,L}^b\right)^{-1} \mathbf{w}_{j-1,L}^{(m)}\right)
\end{aligned}$$

obtained from the last step, the $\psi_m^b(x_{j,L})$ in Eq. (10) can be computed recursively for $j = 2, \ldots, N_T$ similar to Eq. (14) as in Eq. (16).

It is easy to obtain the recursive updating of $\psi_m^b(x_{j,l})$ in Eq.(12) for $l =, L-1, \ldots, 1$ and $j = 1, \ldots, N_T$ as in Eq.(17) with $\psi_m^b\left(x_{0,l}^{(m)}\right) = \psi_m^b\left(x_{N_T,l+1}^{(m)}\right)$.

It is seen that the total complexity of the MSGA-BFS identification step excluding the partial sorting and block operations (for constant channels $\mathbf{H}$ over one block) is $\mathcal{O}(MN_T^2)$ which is the same as that of the MSGA-DFS algorithm.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the fruitful discussions with the researchers at Toshiba Telecommunication Research Lab (Bristol) and the support of its directors. The authors also wish to thank Mr. C. M. Vithanage and Mr. M. Webb for their helpful comments.

REFERENCES

[1] Y. Jia, C. Andrieu, R. J. Piechocki, and M. Sandell, "Gaussian approximation based mixture reduction for near optimal detection for MIMO system," *IEEE Commun. Lett.*, vol. 9, no. 11, pp. 997–999, Nov 2005.

[2] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, pp. 41–59, Autumn 1996.

[3] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short lengths in a lattice, including a complexity analysis," *Mathematics Computation*, vol. 44, no. 3, pp. 463–471, Apr. 1985.

[4] B. M. Hochwald and S. ten. Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[5] D. Pham, J. Luo, K. Pattipati, and P. Willett, "An improved complex sphere decoder for V-BLAST systems," *IEEE Signal Processing Lett.*, vol. 9, no. 11, pp. 748–751, May 2004.

[6] G. Latsoudas and N. Sidiropoulos, "A hybrid probabilistic data association-sphere decoding detector for multiple-input-multiple-output systems," *IEEE Signal Processing Lett.*, vol. 12, pp. 309– 312, Apr. 2005.

[7] J. Jalden and N. Ottersten, "On the complexity of sphere decoder in digital communication," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.

[8] J. Luo, K. Pattipati, P. Willett, and F. Hasegawa, "Near-optimal multiuser detection in synchronous CDMA using probabilistic data association," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 361–363, Sep. 2001.

[9] D. Pham, K. Pattipati, P. Willett, and J. Luo, "A generalized probabilistic data association detector for multiple antenna systems," *IEEE Commun. Lett.*, vol. 8, no. 4, pp. 205–207, Apr. 2004.

[10] Y. Jia, C. M. Vithanage, C. Andrieu, and R. J. Piechocki, "Probabilistic data association for symbol detection in MIMO systems," *IEEE Electron. Lett.*, vol. 42, no. 1, pp. 38–40, Jan. 2006.

[11] J. C. Fricke, M. Sandell, J. Mietzner, and P. A. Hoeher, "Impact of the Gaussian approximation on the performance of the probabilitistic data association with MIMO decoder," *EURASIP J. Wireless Commun. Networking*, vol. 5, pp. 796–800, 2005.

[12] M. S. Yee, "Max-log-MAP sphere decoder," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, Mar. 2005, pp. 1013–1016.

[13] Y. L. C. de Jong and T. J. Willink, "Iterative tree search detection for MIMO wireless systems," in *Proc. IEEE Semiannual Veh. Technol. Conf.*, Fall 2002, pp. 1041–1045.

[14] ——, "Iterative trellis search detection for asynchronous MIMO systems," in *Proc. IEEE Semiannual Veh. Technol. Conf.*, May 2003, pp. 503–507.

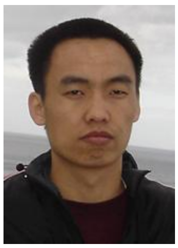
[15] ——, "Iterative tree search detection for MIMO wireless systems," *IEEE Trans. Wireless Commun.*, vol. 53, no. 6, pp. 930–935, June 2005.

[16] D. Guo, X. Wang, and R. Chen, "Multilevel mixture Kalman filter," *EURASIP J. Applied Signal Processing, Special Issue Particle Filtering Signal Processing*, vol. 15, pp. 2255–2266, Nov. 2004.

[17] J. B. Anderson and S. Mohan, "Sequential coding algorithms: A survey and cost analysis," *IEEE Trans. Commun.*, vol. 2, no. 2, pp. 169–176, Feb. 1984.

[18] D. Knuth, *The Art of Computer Programming, Vol. III: Sorting and Searching*. Reading, MA: Addison-Wesley, 1973.

[19] Y. Jia, "Stochastics approximations for reduced complexity signal processing algorithms in MIMO wireless communications," Ph.D. diss., University of Bristol, Bristol, UK, Feb 2007.



Yugang Jia was born in P.R. China in 1979. He received the M.S. degree from Northwestern Polytechnic University, Xi'an, P.R. China in 2000, and the Ph.D. degree in electrical engineering from the University of Bristol, Bristol, UK, in 2007.

He is currently a research scientist with Philips Research Asia - Shanghai. His research interests include statistical signal processing and future wireless communication systems.

$$
\begin{aligned}
\psi_m^b(x_{j,L}) &= \exp\Big(-\left(\mathbf{w}_{j-1,L}^{(m)} - x_{j,L}\mathbf{e}_j\right)^H \left(\mathbf{\Pi}_{j,L}^b\right)^{-1}\left(\mathbf{w}_{j-1,L}^{(m)} - x_{j,L}\mathbf{e}_j\right)\Big)p(x_{j,L})\prod_{k=1}^{j-1} p\left(x_{k,L}^{(m)}\right) \\
&= \psi_m^b\left(x_{j-1,L}^{(m)}\right)\exp\Big(-\zeta_{j,L}^b\left|\varepsilon_{j,L}^{(m)}\right|^2 + 2\Re\left(x_{j,L}\varepsilon_{j,L}^{(m)}\right) - \left|x_{j,L}\right|^2[\left(\mathbf{\Pi}_{j,L}^b\right)^{-1}]_{(j,j)}\Big)p(x_{j,L}), \qquad (16)
\end{aligned}
$$

$$
\varepsilon_{j,L}^{(m)} = \left(\mathbf{w}_{j-1,L}^{(m)}\right)^H [(\mathbf{\Pi}_{j,L}^b)^{-1}]_{(:,j)}, \zeta_{j,L}^b = \left((\gamma-\gamma_l)^{-1} + [(\mathbf{\Pi}_{j,L}^b)^{-1}]_{(j,j)}\right)^{-1}.
$$

$$
\begin{aligned}
\psi_m^b(x_{j,l}) &= \exp\Big(-\left(\mathbf{w}_{j-1,l}^{(m)} - \left(x_{j,l} - x_{j,l+1}^{(m)}\right)\mathbf{e}_j\right)^H \left(\mathbf{\Pi}_{j,l}^b\right)^{-1}\left(\mathbf{w}_{j-1,l}^{(m)} - \left(x_{j,l} - x_{j,l+1}^{(m)}\right)\mathbf{e}_j\right)\Big) \\
&\quad p(x_{j,l})\prod_{k=1}^{j-1} p\left(x_{k,l}^{(m)}\right)\prod_{k=j+1}^{N_T} p\left(x_{k,l+1}^{(m)}\right) \\
&= \psi_m^b\left(x_{j-1,l}^{(m)}\right)p(x_{j,l})/p(x_{j,l+1}^{(m)})\exp\Big(-\zeta_{j,l}^b\left|\varepsilon_{j,l}^{(m)}\right|^2 + 2\Re\left(\left(x_{j,l} - x_{j,l+1}^{(m)}\right)\varepsilon_{j,L}^{(m)}\right) \\
&\quad -\left|\left(x_{j,l} - x_{j,l+1}^{(m)}\right)\right|^2[\left(\mathbf{\Pi}_{j,L}^b\right)^{-1}]_{(j,j)}\Big), \qquad (17)
\end{aligned}
$$

$$
\varepsilon_{j,l}^{(m)} = \left(\mathbf{w}_{j-1,l}^{(m)}\right)^H [(\mathbf{\Pi}_{j,l}^b)^{-1}]_{(:,j)},
$$

$$
\zeta_{j,l}^b = \left((\gamma_{l+1}-\gamma_l)^{-1} + [(\mathbf{\Pi}_{j,l}^b)^{-1}]_{(j,j)}\right)^{-1}
$$

Christophe Andrieu was born in France in 1968. He received the M.S. degree from the Institut National des Telecommunications, Paris, France, in 1993, and the D.E.A. and Ph.D. degrees from the University of ParisXV, Cergy Pontoise, France, in 1994 and 1998, respectively.

From 1998 until 2000, he was a Research Associate with the Signal Processing Group, Cambridge University, Cambridge, U.K., and a College Lecturer at Churchill College, Cambridge. Since 2001, he has been a Lecturer of statistics at the Department of Mathematics, University of Bristol, Bristol, U.K. His research interests include Bayesian estimation, model selection, Markov chain Monte Carlo methods, sequential Monte Carlo methods (particle filter), stochastic algorithms for optimization with applications to the identification of hidden Markov models, spectral analysis, speech enhancement, source separation, neural networks, communications, and nuclear science, among others.

Robert J. Piechocki (M'01) received his M.Sc. degree in electrical engineering from the Technical University of Wroclaw, Wroclaw, Poland, in 1997, and the Ph.D. degree in electrical engineering from the University of Bristol, Bristol, U.K., in 2002. He is currently a Research Fellow at the Centre for Communications Research, University of Bristol. His research interests lie in the areas of statistical signal processing for communications, analog VLSI signal processing, and the optimization of wireless systems.

Magnus Sandell (M'92) received the M.Sc. degree in electrical engineering and the Ph.D. degree in signal processing from Luleå University of Technology, Luleå Sweden, in 1990 and 1996, respectively.

He spent six months as a Research Assistant with the Division of Signal Processing at the same university before joining Bell Labs, Lucent Technologies, Swindon, U.K., in 1997. In 2002, he joined Toshiba Research Europe Ltd, Bristol, U.K., where he is Chief Research Fellow. His research interests include signal processing and digital communications theory. Currently, his focus is on multiple-antenna systems and space-time decoding.